Text Mining-Task 3

Xiang Li

12/7/2021

Task 3

I use the tnum ingested to load my book `My Doggie and I` into test2 number space. The name is `BallantyneR/DoggieandI`. Here is the verification.

```
## [1] "BallantyneR/DoggieandI/heading:0001" "BallantyneR/DoggieandI/heading:0002"
## [3] "BallantyneR/DoggieandI/heading:0003" "BallantyneR/DoggieandI/heading:0004"
## [5] "BallantyneR/DoggieandI/heading:0005" "BallantyneR/DoggieandI/heading:0006"
```

Using sentimentr package do sentiment analysis

According to https://www.r-bloggers.com/2020/04/sentiment-analysis-in-r-with-sentimentr-that-handles-negation-valence-shifters/, I have the sentiment flow of the book sentence by sentence.

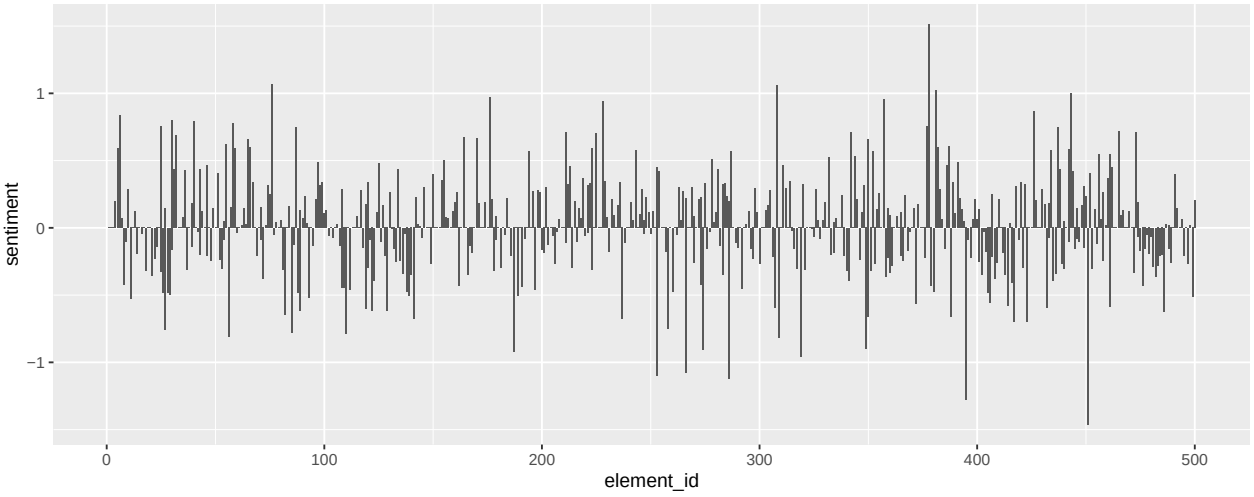
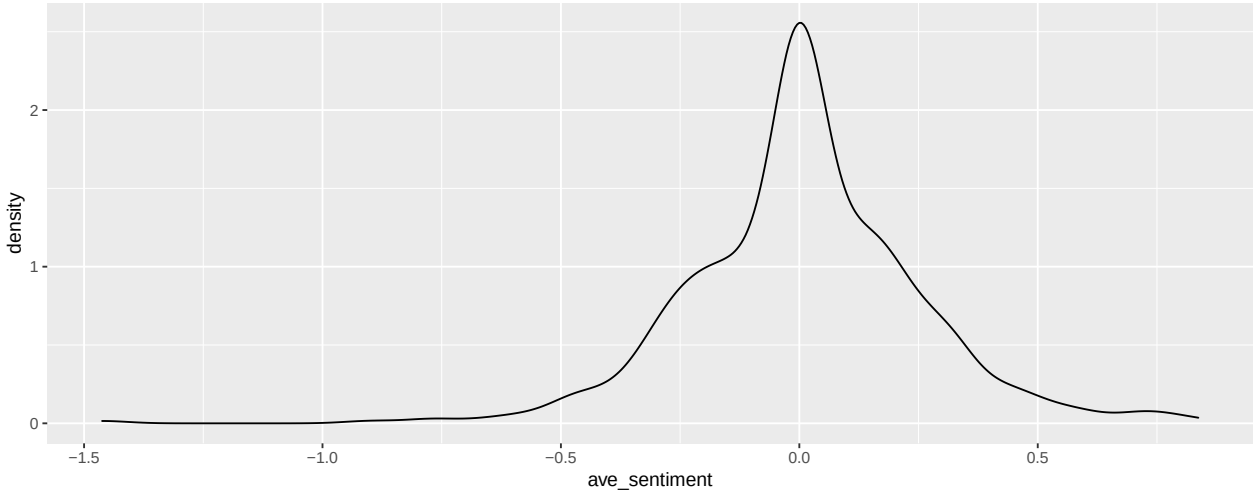


Figure 1: Sentiment flow analysis

Here is the plot of density of average Sentiment of first 500 sentense.



This figure displays the density of average sentiment of the first 500 sentence in `Doggie and I`.

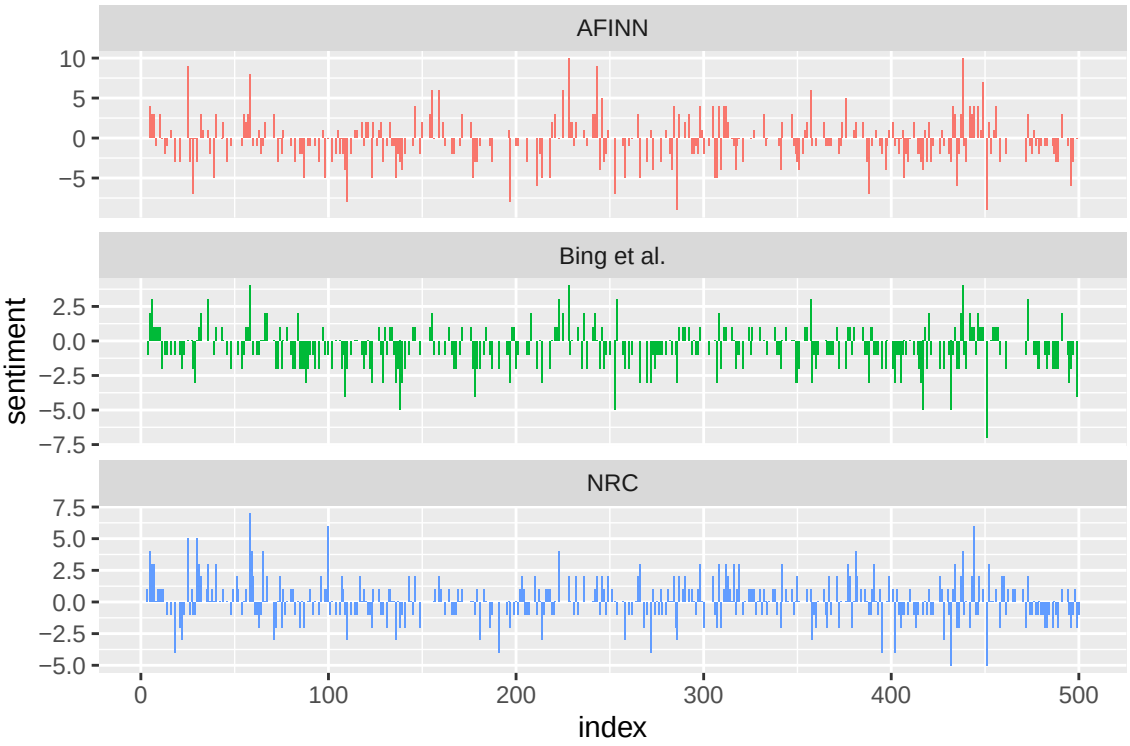


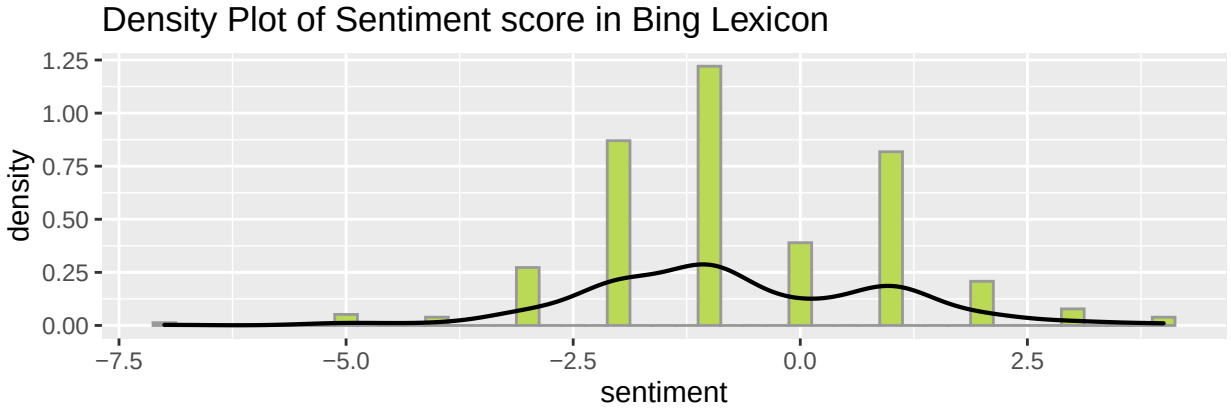
Figure 2: visualization of an estimate of the net sentiment (positive - negative) for each sentiment lexicon

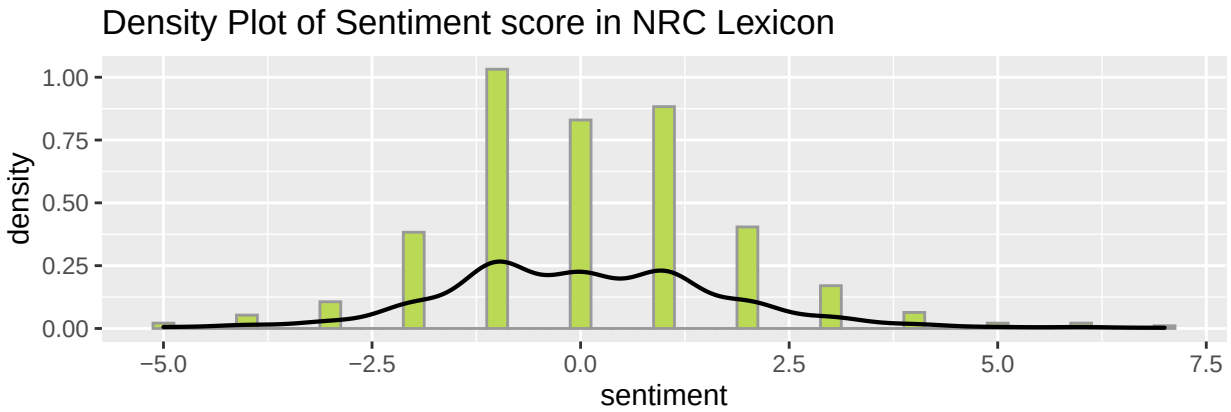
Discussion: Similarly to the task2, instead of using integer division to define larger sections of text every 20 line in the book, I use sentence level in task 3. I have three distinct lexicons for calculating sentiment flows, which produce findings that differ in absolute terms but follow similar relative paths across the novel. In the novel, I detect comparable drops and peaks in emotion at roughly the same locations, but the absolute numbers are much different. The `AFINN` lexicon has the highest absolute values and the highest positive values. `Bing et al` lexicon has lower absolute values and appears to mark longer chunks of continuous positive or negative text. The `NRC` findings are skewed upward in comparison to the other two, positively labeling the text, but identifies identical relative changes in the text.

Here is the summary of affin lexicon

```
##      index          sentiment          method
##  Min.   :  5.0   Min.   :-9.0000   Length:281
##  1st Qu.:123.0   1st Qu.:-2.0000   Class :character
##  Median :266.0   Median :-1.0000   Mode  :character
##  Mean   :256.9   Mean   :-0.4199
##  3rd Qu.:394.0   3rd Qu.: 2.0000
##  Max.   :499.0   Max.   :10.0000
```

```
##     method             index          negative       positive
##  Length:308         Min.   :  4.0   Min.   :0.000   Min.   :0.0000
##  Class :character   1st Qu.:119.8   1st Qu.:1.000   1st Qu.:0.0000
##  Mode  :character   Median :253.5   Median :1.000   Median :1.0000
##                     Mean   :251.0   Mean   :1.481   Mean   :0.8149
##                     3rd Qu.:381.2   3rd Qu.:2.000   3rd Qu.:1.0000
##                     Max.   :499.0   Max.   :7.000   Max.   :5.0000
##    sentiment
##  Min.   :-7.0000
##  1st Qu.:-2.0000
##  Median :-1.0000
##  Mean   :-0.6656
##  3rd Qu.: 1.0000
##  Max.   : 4.0000
```





Extra Credit

character tagging

```
##                                                                 subject property
## 1 ballantyner/doggieandi/section:0002/paragraph:0077/sentence:0001     text
## 2 ballantyner/doggieandi/section:0002/paragraph:0083/sentence:0001     text
## 3 ballantyner/doggieandi/section:0003/paragraph:0001/sentence:0001     text
## 4 ballantyner/doggieandi/section:0003/paragraph:0003/sentence:0001     text
## 5 ballantyner/doggieandi/section:0003/paragraph:0005/sentence:0003     text
## 6 ballantyner/doggieandi/section:0003/paragraph:0013/sentence:0001     text
##
## 1 "Then he began to make a noise v'en I left him; so, bein' sure the bobbies would rout 'im out at la
## 2
## 3
## 4
## 5
## 6
##   numeric.value error unit
## 1            NA    NA   NA
## 2            NA    NA   NA
## 3            NA    NA   NA
## 4            NA    NA   NA
## 5            NA    NA   NA
## 6            NA    NA   NA
##                                                                                                tags
## 1 Willis,Willis,Willis,reference:Willis,reference:Willis,reference:Willis,reference:Willis
## 2 Willis,Willis,Willis,reference:Willis,reference:Willis,reference:Willis,reference:Willis
## 3 Willis,Willis,Willis,reference:Willis,reference:Willis,reference:Willis,reference:Willis
## 4 Willis,Willis,Willis,reference:Willis,reference:Willis,reference:Willis,reference:Willis
## 5 Willis,Willis,Willis,reference:Willis,reference:Willis,reference:Willis,reference:Willis
## 6 Willis,Willis,Willis,reference:Willis,reference:Willis,reference:Willis,reference:Willis
##         date                                  guid
## 1 2021-12-07 ef959883-ebcc-43f4-b85c-6c39ae260a78
## 2 2021-12-07 4b4d11f1-c851-44cb-a8b7-bd602f62a1d1
## 3 2021-12-07 58bae6c7-c301-467c-82ba-48e1245dc9ed
## 4 2021-12-07 475282bf-47f9-4ded-9b99-97b9b76e0e6d
## 5 2021-12-07 c93760a5-1d40-429c-946a-0cffb5997cb3
## 6 2021-12-07 f13d922d-ba42-4061-abd6-70e6c8854fcc
```